Biomedical Data Mining

Extended Abstract

Glycopeptides by quantum chemistry and artificial intelligence

Mehmet Gokhan Habiboglu

Abstract

Glycation destroys or impairs the biological function of peptides and proteins. The bacteria cell wall polymers consist of GlcNAc, which is cross-linked with oligopeptides. While glutamine is a nonessential amino acid that can be derived from glucose, some cancer cells primarily depend on glutamine for their growth, proliferation, and survival. Numerous types of cancer also depend on asparagine for cell proliferation. Thus, glucose and asparagine interactions are at the center of cancer. Moreover, Semliki Forest virus grown in mosquito cells consist of asparagine-linked oligosaccharides. Dengue virus envelope protein (E) consists of two N-linked glyscosylation sites asparagine-67 and asparagine-153. N-linked oligosaccharide side chains on flavivirus E proteins have been associated with viral morphogenesis, infectivity, and tropism. Virologically, ZIKV consists of a single-stranded, positive-sense RNA while the genome encodes three structural proteins including an E protein. Both cryo-electron microscopy and crystallization measurements supported the role of asparagine as a glycosylation site for host cell attachment. Recent studies have shown that ZIKV attacks parts of the adult brain that are central to learning and memory. Furthermore, an observable change in the brain is impaired glucose metabolism within Alzheimer´s disease progression, assessed using positron emission tomography to monitor {18F}-2-deoxy-2-fluoro-glucose uptake within the brain of Alzheimer´s disease patients. The exact molecular mechanism between O-GlcNAc and Aβ remains elusive at the atomic level. Here, we present the structures and energetics of Glc-Asn and GlcNAc-Asn complexes in an aqueous solution medium at the electronic level using quantum chemical calculations linked with artificial intelligence studies. To the best of our knowledge, this study represents the first investigations of aqueous glycopeptides usng quantum chemistry associated with artificial intelligence.

Protein glycosylation is one of the most common post-translational modifications related to protein structure, stability, trafficking, and proteiN-protein interactions. Protein glycosylation is divided into O- or N-glycosylation according to the amino acid binding groups, which include the hydroxyl side chains of serine (S) or threonine (T) and the carboxy-amido nitrogen of asparagine (N) residues, respectively. The heterogeneity and complexity of N-glycosylation are due to the various combinations of four kinds of carbohydrate blocks, including N-acetylhexosamine (HexNAc; e.g., N-acetylglucosamine, N-acetylgalactosamine), hexose (Hex; e.g., glucose, galactose, mannose), fucose (Fuc), and sialic acid (Sia; N-acetylneuraminic acid). These combinations are made by their corresponding glycosyltransferases in the endoplasmic reticulum and Golgi apparatus. Various diseases, including cancer, involve the fucosylation of human N-glycosylation. This is due to two kinds of fucosyltransferase: alpha-1,3/4 fucosyltransferase (FUT 3, 4, 5, 6, 7, 9, 10, and 11) and alpha-1,6 fucosyltransferase (FUT 8)3. The former synthesizes a Lewis or sialyl Lewis structure on the “outer” arm of N-glycan, and the latter produces trimannosyl “core” fucosylation, catalyzing a fucose to the innermost GlcNAc. These fucosyltransferases are highly expressed in cancers, including liver, breast, prostate, noN-small cell lung, and melanoma cancers.

Fucosylation of N-linked glycoproteins can lead to alterations in protein activity in inflammation, immune responses, and cancer metastasis. Core fucosylation has been known as an important key for structural stability and function of N-glycoproteins8. For example, core fucosylation deficient IgG has been reported to may lead to antibody-mediated cellular cytotoxicity. In case of core-fucosylated alpha-fetoprotein, exhibiting an increased affinity for the fucose-specific lectin of lens

Mehmet Gokhan Habiboglu
Turkish-German University, Turkey, E-mail: habiboglu@tau.edu.tr

culinaris agglutinin (LCA), is well-known a biomarker for HCC11. FUT8-mediated alpha-1,6 core fucosylation which interrupts the proteolytic cleavage of L1CAM protein by plasmin, plays a molecular driver of metastasis in melanoma. On the other hand, detailed study of biological role of outer fucosylation in human N-glycoprotein remains uncertain. In mammals, alpha-1,3-outer fucosylated glycans of Schistosoma mansoni and H. pylori are involved in host cell adhesion. For example, in human N-glycoproteins, alpha-1 antitrypsin significantly increases fucosylation in emphysematous lung disease, thereby now it is necessary to study outer fucosylation in detail. Haptoglobin is also decorated with outer fucosylation in pancreatic and gastric cancer. From these studies of N-glycoproteins with various diseases, it is important to identify the detailed structure of core and outer fucosylation.

Recently, liquid chromatography-tandem mass spectrometry (LC-MS/MS) has emerged as a powerful technique for glycoprotein identification. Using tryptic digestion of proteins and tandem MS, we could automatically predict N-glycosylation sites and their attached glycan composition. In collisioN-induced dissociation (CID) spectra from LC-MS/MS analysis, features from several fragmentation ions from N-glycopeptides could be used to determine the type of fucosylation. Glycan fragment ions (B ions), such as Hex-HexNAc (m/z 366.1), Hex-HexNAc-Fuc (m/z 512.2), Sia-Hex-HexNAc (m/z 657.2), and Sia-Hex-HexNAc-Fuc (m/z 803.3), have been used to identify the fucosylation of N-glycopeptides from haptoglobin, hemopexin, complement factor H and kininogen. In addition, N-glycopeptide fragment ions (Y ions) with Fuc and their neutral loss provide additional information regarding the glycan composition within immunoglobulin gamma (IgG). Using manual annotation with B and Y ions from the CID spectra of N-glycopeptides, we successfully identified 71 fucosylated N-glycopeptides from human plasma glycoproteins, e.g., vitronectin, alpha-1-acid glycoprotein (AGP), and IgG; however, the classification of fucosylation has not been performed. Recently, a total of 973 fucosylated N-glycopeptides were identified from prostate cancer cell lines to indirectly determine the fucosylation type using multiple lectin enrichment and LC-MS/MS26. However, there is no software that automatically classifies one of the four fucosylation types as ‘none’, ‘core’, ‘outer’, or ‘dual’ from N-glycopeptides.

The deep neural network (DNN) and support vector machine (SVM), which has mainly been used for supervised machine learning, has advantages of simplicity in generating learning models without overfitting problems. The DNN has recently been used in various fields, including the prediction of gene expression levels in epigenetic models, the sensitivity of molecules, the structure and activity of drugs, the sequence of peptides, and biological images from microscopy, magnetic resonance imaging, and mass spectrometry. However, there are no reports of using DNN methods to predict or classify the molecular structure using peak m/z and intensity values from mass spectrometry, except for an in silico algorithm that predicts the charge and structure of 94 lipid metabolites using CID tandem mass spectrometry. Using the SVM, plasma proteins have been predicted as biomarkers of inflammation with 77% accuracy. Theodoratou and her colleagues showed that the SVM could be applied to classify different glycosylation types of plasma IgG in colorectal cancer prognosis. These reports showed that SVM could be used as a classifier in the bioinformatics fields, such as proteomics and glycoproteomics.

Here, we used MS/MS combined with machine learning methods (such as the DNN and SVM) to classify the fucosylation of N-glycopeptides. The identified N-glycopeptides from IgG and AGP were used for training and testing the machine learning models. Models with the best performance from the machine learning methods were applied to classify unknown fucosylated N-glycoproteins in human plasma.

This work is partly presented at 24th World Chemistry & Systems Biology Conference on October 03-04, 2018 in Los Angeles, USA
October 03-04, 2018 | Los Angeles, USA